



RepoMMan Project

R-D12

Report on admin and teaching & learning user
requirements interview data

Richard Green and Chris Awre

January 2007



The RepoMMan Project

Project Director:	Ian Dolphin, Head of e-Strategy, University of Hull (i.dolphin@hull.ac.uk)
Project Manager:	Richard Green (r.green@hull.ac.uk)
Technical Lead:	Robert Sherratt (r.sherratt@hull.ac.uk)
Repository Domain Specialist:	Chris Awre (c.awre@hull.ac.uk)

The Repository Metadata and Management Project (RepoMMan) at the University of Hull is funded by the JISC Digital Repositories Programme. The project is being carried out by the University's e-Services Integration Group (e-SIG) within Academic Services.

Introduction

The RepoMMan Project Plan requires the development of a front-end interface to the Fedora repository software for users working on their own behalf or collaboratively. Clearly, it is necessary that the development team understand what it is that potential users “do” in their day-to-day work in order to understand how a repository might help, and what their user needs might be. This report deals with investigative work undertaken with potential staff users who have administration roles and/or are members of the University of Hull’s teaching and learning (T&L) community.

Methodology

Similar investigative work with researchers^{1,2} was informed by a two-pronged investigation of research methods. The first approach was to interview a small number of University of Hull researchers at length about their working practices, the second approach was to ask similar questions of a wider (inter)national audience using an on-line survey. From the data thus collected generalised scenarios and use cases were developed. This current document is based on a number of interviews and the findings from the T&L work are supported by a survey conducted in conjunction with the CD-LOR Project.³ It was felt that an on-line survey of administrators would be of limited use at this stage in the project’s work and the decision was taken not to undertake such a process; this is in line with the Project Plan which states that such a survey would only be undertaken “if needed”.

The interviews

Twelve staff from the University of Hull were interviewed at length, taking between 50 and 100 minutes each. The staff were involved in administration, T&L, or both. The outline structure of the interviews sought to elicit information about the digital material they currently produce, the workflow through which they produce it, the specific file types of digital material produced and hence the contribution that a repository facility might make to its development, exposure and - potentially - preservation. Some of those interviewed already had a sound knowledge of what a repository might provide for them, others had little or no knowledge of repositories.

Findings from the interviews

Unlike the process undertaken with researchers, there was no expectation from these interviews that we should be able to draw out a generalised process that described the work these staff do. Rather our intention was to draw out pointers that would allow us to design a repository, and hence the RepoMMan workflow tool, which would support the needs of those that we interviewed and, by extension, their colleagues. There was an assumption that we should provide for previously established researchers’ needs but that we needed to establish and service any additional needs or processes that administrators or those in T&L may have.

¹ Green, R (2006) *R-D4 Report on research user requirements interview data* RepoMMan Project, University of Hull at: <http://www.hull.ac.uk/esig/repomman/documents>

² Green, R (2006) *R-D3 Report on research user requirements ion-line survey* RepoMMan Project, University of Hull at: <http://www.hull.ac.uk/esig/repomman/documents>

³ Margaryan A (2006) *Report on Personal Resource Management Strategies* CD-LOR Project, Glasgow Caledonian University

Teaching and Learning

The T&L interviews ranged over myriad possible requirements for digital objects in the repository. Items not previously raised by researchers included, in approximately increasing order of potential file size:

- laboratory scripts and other notes to assist students (probably Word or pdf files)
- manufacturers' data sheets (for electronic or mechanical components) (probably pdf)
- scripts and files related to the design and operation of a computer or machine (C programming language, Java, Matlab [.m], MathCAD [.mcd], circuit board design [.emx] etc)
- reports (from reports of a single experiment through to theses, probably Word files)
- podcasts [.mp3]
- collections of files that together comprised a learning object (perhaps including .doc, .swf, .jpg, .mov, .wav etc)
- disk images (of working design implementations, for backup and recovery) [.gho and other formats]
- data files [possibly arbitrary file types]
- RAW image files [there is a range of formats described as .raw, many unique to individual camera manufacturers]
- collections of data files comprising a simulation dataset [zip, also need for collections here and especially a need for versioning]
- archives of a complete student learning environment (not just a snapshot of the content, but also the interface and aesthetics)
- large format digital video [.dv]

The first point that should be made is that the files listed here range in potential size from under one kilobyte through to tens of gigabytes: this raises some serious long-term issues for storage. Two further points should be noted: that some of these, quite complex, objects should be searchable for finer-grained, internal, information (eg learning objects, video) and that some might require access by a user group (eg students taking a module as opposed to an owner or public access). There are implications, then, for metadata and access provision.

As was found in interviews with researchers, members of the T&L community were immediately interested in two possible benefits of using a repository for development work: the fact that the repository would be subject to a regular, thorough, backup routine, and the promise of flexible access to material in the repository 'from anywhere'. T&L staff identified two particular benefits of remote access. The first was that it would be very useful to have access to their store of materials when away from the University, perhaps at a conference or meeting in the UK or abroad - this in addition to the benefits of being able to access materials from home. It was suggested that there were many occasions when a member of staff would have liked to have shown something to a colleague at a meeting but didn't have it available on a laptop or via web access; the ability to search for it and access it quickly was seen to be potentially helpful. The second major benefit that was suggested related specifically to the teaching environment. It was suggested that when a student raised something 'unexpected' in a lecture there would be considerable benefit in being able to say (eg) "I have a short video clip about that," that could quickly be searched for, retrieved and shown using web access. A third benefit was mentioned although only by one pair of staff: the possibility of a group of staff (a department, say) being able to share, comment on and refine a pool of teaching materials to which each had contributed.

Administration

The interviews with staff involved in administration at one level or another did not really raise any new issues in terms of file storage, but they did raise some interesting problems for the structuring of digital objects, their associated files and the management of access to them.

In looking to a potential repository as supporting, in part, the official University archive we interviewed a member of Committee Section, the group which services the committees which make up the legal structure of the University. She identified some interesting needs related to the development of a document as it passed through a number of stages traversing up through a sequence of hierarchical committees, not least the need to have each committee's version of the document preserved (and clearly identifiable) for posterity. This need was echoed by a member of staff from the University Quality Office, a team of staff who work quite closely with Committee Section.

Administrators, by the nature of their work, are often involved in developing a document (a report, say) from a draft through to a published item. Along the way, they have to collate and synthesise comments and suggestions made by contributors to the process. A number of those that we interviewed had ideas about how a repository could potentially help with this process. In particular, some of them were enthusiastic about the idea of making a draft available to a small group of people for comment through the repository with the comments being fed back into the digital object itself; more than one way of implementing this was suggested.

Administrator interviewees fell into two camps when it came to the question of backup: some were largely uninterested in the issues because the nature of University provision for their workplace meant that backup was centrally managed, automated and regular, the second group did not have such automated provision and were thus attracted to what the repository might offer them in this respect.

A number of our interviewees told us that they regularly took work home and that the facility a repository might provide for remote access would be useful. There is a crossover here with the enthusiasm expressed by a number of them for a repository's help with versioning.

In general, the administrative role generated more interest than T&L in the whole area of metadata (with one notable exception). In describing their administrative work, a number of interviewees reported that they often found themselves looking for 'obscure' documents on the web and they could therefore see the benefits of good metadata and justify any reasonable effort put into generating it.

Scenarios

From the interviews, it is possible to draw out a number of scenarios that encapsulate potential user needs. There is not necessarily a one-to-one correspondence with the interviews that we held with practitioners; in some cases we have associated one person's requirements with a different interview or omitted requirements from one scenario that are adequately dealt with in another. Some interviews have been conflated.

Teaching and learning

Although these scenarios are drawn from interviews with members of the teaching and learning community, each of those staff interviewed had administrative responsibilities of one sort or another. Accordingly these scenarios are not wholly about T&L issues.

Scenario 1

Peter is interested in making available learning objects consisting of between five and nine 'information objects', as he calls them, which each represent about five minutes of T&L activity. The learning objects would be searchable for their information components, and these information components would be capable of extraction and reconstruction into new learning objects. An information object might consist of a brief introduction, some facts that are being taught, and an assessment. It may also contain media of some sort, perhaps an image, a Flash object, or a short video clip. Peter also sees advantages of having aggregations

of learning objects stored in his private user area - each perhaps corresponding to a lecture of some sort. In saying this he is thinking beyond lectures for full-time students and thinking about his need to be able to mount a lecture or short course for an external group at short notice.

Stephen, who works with Peter, likes these ideas and sees much potential advantage in being able to identify components of the learning objects, being able to extract them, and then being able to recombine them, perhaps with additional material. He sees benefits not only in re-using Peter's 'information objects' but in being able to extract components from them - perhaps just the image or video clip. He also sees benefits in students having access to all these components so that they can directly use part of the learning material, perhaps in the assessment work that they do in response to it.

Stephen would like to have all the components of his teaching materials available to him online in a repository. This would enable him to locate a specific item in his private repository space very quickly, to the extent that it might be feasible to locate it in response to an unexpected question from a student in a lecture. He would greatly value the ability to very quickly find and display a short video-clip, say, in this manner.

Stephen can also see considerable value in sharing his materials with others and in users being able to annotate the sub-components. Such comments might say that "this image was really useful in trying to explain X to my students today": in other words offering to others the context in which it was found useful. In sharing objects, Stephen would like the description available to others to be very broad so that an Economist (say) might be able to locate, recognise and use a Flash animation that he had produced for his subject, which is actually Geography. He could also see much benefit in being able to maintain a 'wish list' (à la Amazon) for components that he would like but has not yet found. The list would clearly need to be visible to others.

Scenario 2

Keith works with students who produce a lot of media material: images, video and some sound. At present the finished items reside on a server in his department, which is to say that they are not available for wider use. He regards this as a shame because he and his team are careful about copyright and other permissions so that in most cases the images and clips could be more widely used within the University community if they could be made accessible; a smaller number of these materials could be made available to the public. However, he accepts that this would not be a trivial task because whilst many of the materials have some metadata associated with them this is in a proprietary database at present and it is not clear whether this could be transferred to a repository in an automated way. The materials vary in file size from digital video (dv) format files measured in gigabytes to podcasts and some images which may be only a few megabytes.

Keith can see the future benefit in using a repository such as the University proposes. Students could place their finished materials in the repository and provide appropriate metatagging. He is very interested in the ideas of being able to provide metadata for video that references scenes in the clip by timecode and of being able to annotate images or video. He is not quite so interested in having students use the repository as a development space because he encourages the use of 'social' storage on the web where the drafts can be widely shared with others for their comment, that said he can see potential uses for the inherent versioning capability of the repository, particularly because he and his colleagues encourage their students to reflect on the various stages of the development process.

Backup is a serious concern to Keith at the moment and the repository could potentially offer him a solution to this problem, however already Keith's materials consume almost two terabytes of storage and this could be an issue.

Scenario 3

Timothy works with students to produce simulations of industrial processes. This involves data modelling leading to simulations, programming, and the construction of bespoke computers to service particular needs.

He and his students have a number of relatively unusual storage requirements. Departmental projects tend to build on what went before, rather than to be parallel developments. This means that in preserving previous material which may be extended in later years it is necessary to preserve all aspects of the hardware and software that have gone into a development: programs, data, component specifications, printed circuit board design, overall design and a complete image of the system's disk(s). In terms of file size for storage, some of these components are relatively trivial, however disk images can amount to tens of gigabytes and the data sets used to develop a simulation can be similarly large. The component parts of the project materials would need to be reliably collected together and simple to retrieve.

Timothy can see use for the repository during the development of, most especially, simulations. The development of the data and the model which provides the simulation can be a long process measured in months. It is not unknown to make a fundamental misjudgement in the development which will eventually mean having to revert to a version from many weeks earlier. Automatic versioning in a repository would be useful here. He can also see use for the repository as a backup for development work in progress. The University does not normally provide him with centralised backup although it does provide some centrally maintained systems - which is to say that it manages the basic computer disk image and desktop. Experience has led him to conclude that he should have his own backups of the computers that his students to develop project work in order to guarantee quick restoration in the case of problems. In this regard he is thinking about putting complete disk images in the repository; these would necessarily be many gigabytes in size, even when compressed.

In his administration role within the department, Timothy can see merit in using a repository to store students' completed work, including perhaps electronically marked materials.

Scenario 4

Pradesh manages a resource centre for academics and their students. The centre offers printed materials, some of which it develops itself, a question bank, images, video and software. Currently this material is made available to users through a number of websites and Pradesh can see benefits in having them available from one. He sees a repository, such as the University envisages, as an appropriate place for these materials for two reasons: the repository might be of use whilst developing some of the, especially printed, materials; and some of the materials produced have long-term value and should continue to be available even should his Centre close.

The Centre produces a number of publications each year. At the moment, these tend to be drafted by one of Pradesh's team but then checked and commented on by a number of others. A private repository area that allowed collaborative working could be useful in this process.

Other materials that the Centre provides are not so much developed by them as by others. However, Pradesh sees the repository potentially as a useful showcase where all the materials can be brought together. He accepts that these materials would need to have good metadata to aid search and discovery. Many of them do have such metadata at the moment, but it is not directly 'attached' to the items, rather it is in an associated database. The question bank poses a slightly more complex problem in that access to it, should it be done through a repository, would have to be subject to flexible yet absolute security.

Pradesh understands the copyright issues that would be involved in providing his Centre's materials through a repository and is in a position to address them.

Administration

As with the T&L scenarios, where student contact is involved there is here sometimes a level of crossover in individual roles between the main area of administration and T&L activities.

Scenario 5

Tabatha works in an academic department with students, but her role centres significantly on administrative tasks related to the courses that the department provides. She organises programmes of study which normally involve students spending significant time in the workplace, and she is involved in the quality assurance of these courses - something which is a particular concern given the level of external contribution from the workplace teams.

A large part of Tabatha's work is the bi-annual Quality Assurance Report that she writes. This involves gathering together a significant volume of information but then she writes the document without collaboration. This is a long process which often involves working at home or whilst on the move. Tabatha is interested in the idea of using a repository to store this work-in-progress so that versioning is done for her, the document is available to her via the web, and so that automatic, routine backup of the developing draft takes place.

Tabatha's department has to deal with a range of written assignments, dissertations and theses. At the moment many of these are, by statute, paper documents which are marked by more than one assessor. Following graduation many of these documents might usefully be referenced by others but at present they are stored in a somewhat obscure office. Tabatha can see that allowing students to submit these materials in electronic form might have a number of advantages: the electronic form could be made easily available to multiple markers; once marked, the document could be made more widely available for future reference in a repository; and the document could be preserved over time.

Scenario 6

Mary works within the University committee structure. She is responsible for providing agendas, minutes and working papers to committee members and for making them available to other interested parties.

Although the team in which she works is given shared facilities on the University network, which also provides backup, she would like to have in place more flexible sharing arrangements which encompass, not just the team with which she works, but authorised members of University staff. She can see how a repository might help with this, and also how it might address some of her longer term access and preservation needs. (Some of the documents she deals with are the official records of University business and need to be preserved indefinitely.)

Mary may be provided with a report which is to go to a sub-sub-committee. This arrives from its sponsor in a finished, digital form and she associates it with an agenda and other papers for an upcoming committee meeting. (There will probably be a set of previous meeting minutes for the committee too.) The report must be clearly identified as the report that went to this committee. Mary would like to be able to make these pre-meeting papers available in digital form for members of the committee to access, and for the information of a small group of other legitimately interested parties.

Following the meeting, the report may be accepted, sent back to its sponsor for revision and resubmission or passed up to the next committee level. In these last two cases, the cycle starts again in that the (possibly) revised paper goes into the collection for the next committee meeting in a version that must be clearly associated with the new meeting (and not the previous one). If the report is going to a new committee the new document needs appropriate access permissions for its members. It is possible that the report would work its way up to the

highest committee for final approval and this would result in, say, five possibly distinct versions each of which needs to be clearly identified, associated with a particular committee meeting and preserved. At whatever level the report is signed off, the final version may need to be made available to a much wider audience.

The agendas and minutes from all these meetings are normally made available to University staff after the event, with the exception that business is divided into Part A and Part B business: the agendas for both parts are published, but only the minutes and associated papers from Part A are published after the meeting. Part B materials contain confidential information, some of which may always be so, other of which may be made public after some time. The status of any such confidential material would be reviewed each time a request was made to see it.

The security of papers stored in a repository pre-meeting would rely on the University Computing Service maintaining up-to-date lists of committee members for access by other University systems.

Scenario 7

Lauren works with a team responsible for codes of practice within the University. Although the team in which she works is given sharing facilities on the University network, which also provides backup, she would like to have in place more flexible sharing arrangements which encompass, not just the team with which she works, but staff who contribute to the development of and compliance with the codes; these staff are not necessarily all at the University of Hull but may be employees of a partner organisation. She can see how a repository might help with this.

Lauren and her colleagues develop a new document from version 0.1 onwards. Each sub-version is retained. At stages in the development process she would like to make the drafts available to others outside her department for comment. She would like to do this in a linear fashion so that the second contributor can see the comments made by the first, and so on. This is an iterative process which eventually becomes version 1.0 which is then made available to members of staff at the University and in partner institutions. In due course the document will come up for review at which time version 1.0 may be subject to minor modification and become 1.1 (say), or it may need major revision in which case a new iterative process will start with a view to producing version 2.0 and this process will probably use 1.0 as its starting point. 1.0 would need to be preserved for the historical record. Lauren can see how the automatic versioning offered by a repository could help with all this and how a collaborative facility could be used effectively.

The documents produced in this way are complex. Lauren would like to see them exposed to their end-users as a collection that is easily and effectively searchable so that relevant documents and sections can be located quickly.

Lauren and her colleagues in the team are also responsible for monitoring compliance with the codes. This involves to-and-fro work with departments on documents and Lauren would like to have access to some of these works-in-progress before their formal submission without having to keep asking the department for the latest version.

Scenario 8

Julia is an administrator whose job involves reporting each year on the teaching in her area of the University. This report is made against the background of Codes of Conduct, course specifications and other internal systems.

In order to prepare the report each year, Julia and members of the group she services require access to the latest documents describing University practice and requirements. Not all of the

group are members of University staff, some work for partner institutions. At present she and colleagues sometimes find it difficult to locate what they want over the web. Julia hopes that the use of a repository to store these materials would allow quick and effective search and discovery. She understands that effective metadata would be required to achieve this but hopes that the authors of the papers will see the merit in providing it. As a *quid pro quo* she would be willing to spend a small amount of time making sure that her own documents had effective metadata in order to facilitate their use by others.

User needs

The structure of RepoMMan's Project Plan anticipated that the requirements of potential users in the T&L and administration communities would extend the user needs of those in the research community rather than contradict them. This has turned out to be the case. It is therefore appropriate to take the user needs identified in *R-D4 Report on research user requirements interview data* and expand upon them.

- we take in as a *sine qua non* that a repository interface should not make it difficult to do something that is currently achieved easily

Both administrative and T&L interviewees have identified ways in which a repository might make more easy processes that they need to undertake, and make possible some processes that they would like to undertake.

- the repository interface must allow structuring of a user's personal storage space and have the capacity to hold potentially large numbers of objects, possibly of a range of differing types, for each user

The T&L interviews have extended considerably the range of file types that we envisage might eventually be placed in the repository. Additionally, it is clear that the need to cope with large files in a digital object (potentially in multiple versions) is more common than we might have supposed.

The requirements survey in *R-D4* did not specifically address the issues of remote access to these files or backup of them, because at the time we saw this as a service that the repository would offer, rather than a need. It has now become apparent that reliable backup *is* a serious user need at the University of Hull and that remote access would make possible use of some of this 'private' repository material in ways that might not otherwise be feasible.

- the repository should provide an easily usable versioning facility (it must be easy to version a file *and* to revert to an earlier version)

This need has been echoed in these interviews, particularly by people commenting from the standpoint of an administrative role. Some have extended our thinking on the versioning process to include its potential use as part of collaboration techniques.

- the repository should allow sharing of a private document with a closed group of collaborators and should provide some sort of locking facility so that conflicting revisions cannot occur

This need has been reinforced by these interviews. In addition, the interviews have extended our understanding of how a flexible permissions process could be useful. We had thought, previously, that secondary access to a 'private' area would be useful primarily for sharing work in a collaborative process. Our understanding of what might be covered by 'collaborative process' has been extended and, additionally, we have been made aware that read-only access to 'private' repository objects by groups (committee members, a teaching group)

would be very useful. As part of this we acknowledge the need for easily changing the membership of a group, although we do not see this as a function of the repository.

- the repository must make public exposure of content easy and controllable, taking account of digital rights issues as part of that process

This series of interviews has extended our understanding of what is required by way of metadata functionality within the repository. The issue of digital rights is of concern to some in T&L, in administration the concern is more in terms of the Freedom of Information Act and the balance between its requirements and the need for confidentiality of some documents.

